

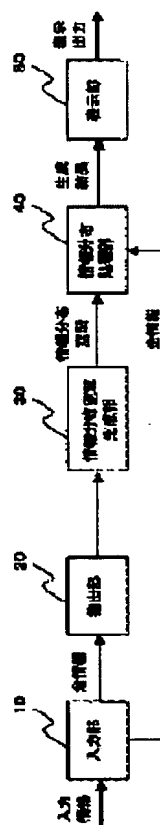
DEVICE AND METHOD FOR INFORMATION DISTRIBUTION

Patent number: JP2002041545
Publication date: 2002-02-08
Inventor: KAKIMOTO TOSHIHIRO
Applicant: FUJITSU LTD
Classification:
 - International: G06F17/30; G06N3/00
 - european:
Application number: JP20000224443 20000725
Priority number(s):

Abstract of JP2002041545

PROBLEM TO BE SOLVED: To provide an information distribution device with superior listing property which efficiently generates and learns an information distribution space, shortens the information distribution processing time, and eliminates the repetitive arrangement of information.

SOLUTION: Information as an object of information distribution processing and a set of its feature vectors are inputted from an input part 10. An extraction part 20 extracts a specific number of pieces of information and feature vectors. Learning is carried out by using feature vectors of information extracted by an information distribution space generation part 30 to generate an information distribution space corresponding to the specific number of pieces of information. An information distribution processing part 40 expands the information distribution space from a specific number of pieces of information to all pieces of input information. The expanded information distribution space is used to distribute all the pieces of input information and obtains an information distribution, which is displayed at a display part 50. When there is an overlap distribution of information, it is rearranged at a nearby unarranged grating by a fixed procedure.



Data supplied from the esp@cenet database - Patent Abstracts of Japan

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-41545
(P2002-41545A)

(43) 公開日 平成14年2月8日 (2002.2.8)

(51) Int.Cl. ⁷	識別記号	F I	テームト* (参考)
G 0 6 F 17/30	2 1 0	G 0 6 F 17/30	2 1 0 D 5 B 0 7 5
	3 5 0		3 5 0 C
	3 6 0		3 6 0 Z
G 0 6 N 3/00	5 6 0	G 0 6 N 3/00	5 6 0 A

審査請求 未請求 請求項の数 5 O L (全 17 頁)

(21) 出願番号 特願2000-224443 (P2000-224443)

(22) 出願日 平成12年7月25日 (2000.7.25)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 柿元 俊博

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74) 代理人 100095555

弁理士 池内 寛幸

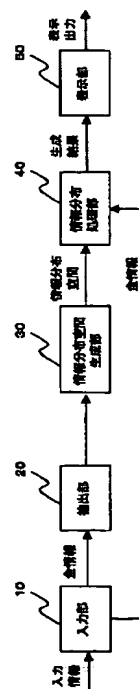
Fターム (参考) 5B075 ND16 NR12 PQ02 PQ13 PQ46
PR06

(54) 【発明の名称】 情報分布装置および情報分布方法

(57) 【要約】

【課題】 情報分布空間の作成学習を効率的に実行し、
情報分布処理時間を短縮し、情報の重複配置を解消し、
一覧性に優れた情報分布装置を提供する。

【解決手段】 入力部10から情報分布処理を実行する
情報とその特徴ベクトルの集合を入力する。抽出部20
により所定数の情報と特徴ベクトルを抽出する。情報分
布空間生成部30において抽出された情報の特徴ベクト
ルを用いて学習して上記所定数の情報に対応する情報分
布空間を生成する。情報分布処理部40において情報分
布空間を所定数の情報から全入力情報に対応するもの
に拡大処理を実行する。拡大後の情報分布空間を用いて
全入力情報を分布し、情報分布を得て、表示部50にお
いて表示する。情報の重複分布がある場合、一定の手順
により近傍の情報未配置格子などに再配置する。



【特許請求の範囲】

【請求項1】 情報の集合とその情報に対応する特徴ベクトルの集合を入力する入力部と、

前記入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出する抽出部と、

前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布する情報分布空間を生成する情報分布空間生成部と、

前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布する情報分布処理部を備えたことを特徴とする情報分布装置。

【請求項2】 前記情報分布空間への特徴ベクトルの類似度を基にした情報の分布処理として、自己組織化マップによる分布処理を用い、

前記情報分布空間生成部が、前記自己組織化マップにおける前記情報分布空間への学習対象情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じて制限する請求項1に記載の情報分布装置。

【請求項3】 前記情報分布処理部が再配置処理部を備え、前記再配置処理部は、前記入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選び、該再配置する情報を該重複配置位置周囲の一定範囲で情報が配置されていない位置に再配置する請求項1に記載の情報分布装置。

【請求項4】 情報の集合とその情報に対応する特徴ベクトルの集合を入力し、

前記入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出し、

前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布する情報分布空間を生成し、

前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布することを特徴とする情報分布方法。

【請求項5】 情報の集合をその特徴ベクトルを基に分布する情報分布装置を構成する処理プログラムを記録したコンピュータ読み取り可能な記録媒体であって、情報の集合とその情報に対応する特徴ベクトルの集合を入力する処理ステップと、

前記入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出する処理ステップと、

前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布する情報分布空間を生成する処理ステップと、

前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布する処理ステップとを備えた処理プログラムを記録したことを特徴とする

記録媒体。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、情報を配置して表示するための情報分布空間作成装置およびその方法に関する。例えば、インターネット上のWWWシステムのブラウジング検索などに利用することができる。

【0002】

【従来の技術】インターネット上のWWWシステムなどを利用した情報化社会の進展によって、マルチメディア情報を用いるアプリケーションとして様々なものが想定され、大量のマルチメディア情報を有効に活用する必要が生じている。これらのマルチメディア情報を活用するためには必要とするマルチメディア情報を的確に検索する必要がある。しかし、従来のキーワード検索に代表されるように、キーワードなど関連付けられた特徴を指定し、マルチメディア情報を絞り込んでゆく技術では、効率性において限界があり、目的とするマルチメディア情報を容易には検索できなかった。主に、検索に利用するキーワードなどの特徴のみでは検索対象であるマルチメディア情報を的確に表現するのが困難であるからである。例えば、キーワード検索において検索対象となるマルチメディア情報に対する適切なキーワードを選択することは非常に難しい問題であり、マルチメディア情報が画像データである場合、特徴指定としてどんなキーワードを使って良いのか分からない場合が多い。また、インターネット上のマルチメディア情報は種々雑多な情報が混合されており、検索結果としてヒットした情報が非常に大量になってしまう場合が多い。これらの問題を解決する手段として、マルチメディア情報を事前に適度な基準を用いて分類しておいて、この分類に従って絞り込んでゆくという分類検索法もあるが、検索対象となるマルチメディア情報は、大量であり、また、動的に日々生成され消滅してゆくことが想定され、分類検索法では対応不可能な場合も多く、その分類も検索目的によりいろいろなものを用意しておく必要があり、現実的ではない。

【0003】これらの状況を解決する手段として、検索結果を情報の特徴により自動分類し、検索結果を配置して表示するための情報分布空間を作成し、その検索結果を情報分布空間に配置し、その中をブラウジング検索する方法がある。このような情報分布空間を作成する手法としては、クラスタリング法やMDS法や自己組織化マップ(SOM)法などを利用したものがある。

【0004】従来の手法の中では、情報分布空間を作成する手法として、多様な情報に対して広く適用できるという観点から自己組織化マップ法を利用したものももっとも適していると考えられている。

【0005】

【発明が解決しようとする課題】しかし、上記情報分布空間を作成する従来のクラスタリング法やMDS法や自

己組織化マップ法などを利用した手法では、以下の問題があった。

【0006】第1の問題点は、情報分布空間を作成し、情報の配置・分布を学習させつつ情報を配置・分布するためには膨大な処理時間がかかることである。コンピュータリソースの向上を考慮しても、なお検索者に対する処理待ち時間が発生し、使い勝手の上で問題があった。上記従来のクラスタリング法やMDS法や自己組織化マップ法などを利用した手法では、情報分布空間の生成処理、情報の配置・分布処理において、情報数の2乗に比例した計算時間を要し、大量の情報の配置・分布には現実的に適用するのが困難であった。一般にブラウジング検索に用いるマルチメディア情報数は大量であり、従来の自己組織化マップ法などの手法では処理時間が大きくなってしまう。例えば、クロック数が500MHz程度のCPUを持つパーソナルコンピュータをリソースとし、1,000件程度の画像を対象とした場合、処理待ち時間は実用的な範囲を超えるほど大きくないが、10,000件程度の画像を対象とした場合、待ち時間が相当長くなってしまい、利用者の使い勝手を大きく劣化させる結果となる。自己組織化マップ法では、自己組織化マップアルゴリズムを用いて、入力特徴ベクトルの配置・分布を学習させて行くが、この学習に多くの時間を要する。

【0007】第2の問題点として、生成した情報分布空間に対し、実際に情報を配置・分布してゆくと、同じ位置に重複配置されてしまう情報が多く、一覽性に欠けるという問題がある。現実には検索対象となる画像などのマルチメディア情報を情報分布空間に配置・分布してゆくと結果として同じ位置に重複配置されてしまう情報が多く、クラスタの大枠を把握するのには問題ないが、実際にブラウジングして一つの情報を検索する場合には却って一覽性に劣り、探索効率が悪くなるという問題が発生する。

【0008】本発明は上記問題点を解決し、情報分布空間の作成、情報の配置・分布の学習を効率的に実行し、情報の配置・分布処理時間を短縮した情報分布装置および方法を提供することを目的とする。

【0009】また、本発明は、生成した情報分布空間における情報の重複配置・分布を効率的に解消し、一覽性に優れた情報分布装置および方法を提供することを目的とする。

【0010】

【課題を解決するための手段】上記目的を達成するため、本発明の情報分布装置は、情報の集合とその情報に対応する特徴ベクトルの集合を入力する入力部と、前記入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出する抽出部と、前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布

する情報分布空間を生成する情報分布空間生成部と、前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布する情報分布処理部を備えたことを特徴とする。

【0011】上記構成により、抽出した所定数の情報を用いて学習して情報分布空間を生成するので、全情報を用いて情報分布空間を生成する場合に比べ、処理時間を効率的に短縮することができ、このように抽出した情報を用いて学習して生成した情報分布空間上に全情報を配置するので、全情報を分布するにあたり、一定品質の適度な分布をもった情報分布処理を実行でき、処理時間の効率化と分布結果の品質維持の両方を同時に実現することができる。なお、生成した情報分布空間に対して全情報を分布して行く時間はその処理が単純なため、全情報を用いて学習して情報分布空間を作成する時間に比べて、少ない時間で実行することができる。

【0012】また、本発明の情報分布装置は、情報分布空間への特徴ベクトルの類似度を基にした情報の分布処理として、自己組織化マップによる分布処理を用い、情報分布空間生成部が、自己組織化マップにおける前記情報分布空間への学習対象情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じて制限することが好ましい。

【0013】上記構成により、情報分布空間の生成において実行する情報の分布最適位置の探索範囲を学習回数に応じて制限することにより学習時間を効率的に制限し、処理時間の効率化と分布結果の品質維持の両方を同時に実現することができる。

【0014】ここで、前記抽出部が、入力された情報の中から所定数の情報を抽出するにあたり、特徴ベクトルが単位ベクトルまたは零ベクトルである情報以外の情報を優先して抽出すれば、各配置位置において適度な特徴ベクトルの分布を持つ情報分布空間を生成することができる。つまり、対象情報の特徴ベクトルとして単位ベクトルまたは零ベクトルが多い場合、ランダムに情報を抽出すると空間の配置が違和感の多いものになる可能性があり、単位ベクトル以外を優先してランダム抽出する方法を取れば違和感のない情報分布空間にすることが出来る。

【0015】また、前記情報分布処理部が再配置処理部を備えることにより、重複配置された情報を見易く配置し直すことができる。ここで、再配置処理部による再配置処理には複数通りの処理がある。入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選ぶ。第1には、この再配置する情報を該重複配置位置周囲の一定範囲で情報が配置されていない位置に再配置する処理がある。つまり、一定範囲に再分布するため、情報配置の違和感を少なくし、かつ、一覽性を上げることができる。さらに空間の空部分の有効利用が

可能になる。第2には、この再配置する情報を該重複配置位置周囲で情報が配置されていない位置であってその位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度が一定範囲にある位置に再配置する処理がある。つまり、情報を空間上の情報分布に忠実に配置することができる。第3には、この再配置する情報を該重複配置位置周囲の一定範囲でかつ情報が配置されていない位置であってその位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度が一定範囲にある位置に再配置する処理がある。第4には、この再配置する情報を該重複配置位置周囲で情報が配置されていない位置であってその位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度と、前記重複配置の位置と前記情報が配置されていない位置との距離との2つのパラメータを持つ関数とし、前記関数値を評価して再配置する位置を決める処理がある。第5には、この再配置する情報を該重複配置位置周囲で情報が配置されている位置を選び、該選択位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度が該選択位置に対する特徴ベクトルと既に該選択位置に配置されている情報が持つ特徴ベクトルとの類似度より大きい場合、該再配置する情報を該選択位置の前面に配置する処理がある。

【0016】本発明の情報分布装置を実現する処理プログラムを記録したコンピュータ読み取り可能な記録媒体を提供すれば、コンピュータ装置を利用して本発明の情報分布装置を構築することができる。

【0017】

【発明の実施の形態】以下、本発明の情報分布装置およびその方法の実施形態について、図面を参照しながら説明する。

【0018】（実施形態1）本発明の情報分布装置は、情報の集合とその情報に対応する特徴ベクトルの集合を入力とし、それら情報が持つ特徴ベクトルの類似度により該情報を情報分布空間に分布・配置して利用者に見せるものであるが、適度な特徴ベクトルの分布を持つ情報分布空間を効率的に短時間で生成する。本発明の情報分布装置は、情報分布空間を効率的に生成するため、入力された全情報のうち所定数の情報を用いた学習により適度な特徴ベクトル分布を持つ情報分布空間を効率的に生成する。本発明の情報分布装置は、一実施形態として、自己組織化マップ（SOM）を用いて情報分布空間を学習して生成するが、処理時間の問題を解決するため、入力情報が多い場合（例えば1,000件以上）には入力の情報集合から抽出した所定数（例えば1,000件程度）の学習対象情報を用いて情報分布空間を作成し、作成された情報分布空間を全入力情報に対応する大きさに拡大し、その情報分布空間を用いて全情報を情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布する手法で対応する。また、自己組織化マップのアルゴリ

ズムを見直し、自己組織化マップにおける情報分布空間への入力情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じて制御し、より学習効率を向上して学習時間を短縮して情報分布空間の効率的生成を行う。さらに、応用機能として、特定の情報集合では特徴ベクトルとして単位ベクトルや零ベクトルが多い場合があるが、このような場合では適切な特徴ベクトルの分布を持つ情報分布空間を生成できないため単位ベクトルや零ベクトル以外の情報ベクトルを持つ情報を優先して抽出してそれら情報を基に情報分布空間を生成する方法を取る。

【0019】図1は、本発明の情報分布装置の構成例を示す図である。図1において、10は入力部、20は抽出部、30は情報分布空間生成部、40は情報分布処理部、50は表示部である。

【0020】入力部10は、情報の集合とその情報に対応する特徴ベクトルの集合を入力する部分である。例えば、情報として画像情報とその画像情報が持つ特徴ベクトルとが入力される。画像情報に対する特徴ベクトルの例としては、画像に付されたキーワード、画像中に写り込んでいるオブジェクト、画像中に写り込んでいるオブジェクトの形状、画像の色要素（色相）、画像の色スペクトルなど多様なものが想定できる。なお、入力部10は画像情報とその特徴ベクトルの入力を受けるという構成でも良く、また、応用機能として、入力部10が、入力された画像から特徴ベクトルを抽出する特徴ベクトル抽出機能を備える構成としても良い。入力部10から入力された全情報と特徴ベクトルは抽出部20および情報分布処理部40に出力される。

【0021】抽出部20は、入力部10から入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出する部分である。また、応用機能として、入力された情報の中から所定数の情報を抽出するにあたり、特徴ベクトルが単位ベクトルまたは零ベクトルである情報以外の情報を優先して抽出する機能を有している。

【0022】全入力情報に対して最適な情報分布空間を生成するには全入力情報を用いて特徴ベクトルを最適に分布させる情報分布空間を生成するが、本発明は、所定数を超える情報入力があった場合、所定数の情報を抽出し、その所定数の情報の範囲で特徴ベクトルを最適に分布させる情報分布空間を生成するものである。このように所定数の情報を用いて情報分布空間を生成しても画像のブラウジング検索などに用いるための適度な一貫性を与えることができる特徴ベクトルの分布を持つ空間とすることができる。なぜなら、表示される情報に対する厳密な意味での最適化された特徴ベクトルの分布を持つ情報分布空間でなくとも、ブラウジング検索に用いるための適度な一貫性と特徴ベクトルに応じた情報の分布が与えられれば十分実用に耐えうるからである。

【0023】情報分布空間生成部30は、抽出部20に

より抽出された情報の集合を用いて学習し、該情報が持つ特徴ベクトルの類似度により該情報を分布する情報分布空間を生成する部分である。例えば、情報分布空間の作成手法としてKohonenの自己組織化マップを使い、生成される情報分布空間は $L \times L$ の正方格子であり、格子点（情報の配置位置）には学習結果として特徴ベクトルが与えられたものとなっている。後述の情報分布処理部40により、この生成された情報分布空間に対して入力された全情報がその特徴ベクトルと格子点の特徴ベクトルとの類似度が最も高い格子点に対して配置される。

【0024】なお、情報分布空間生成部30は、一実施形態として、自己組織化マップによる分布処理を用い、自己組織化マップにおける情報分布空間への入力情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じて制御して学習効率を向上し、短時間に情報分布空間を生成する機能（最適位置探索範囲制御機能）を有する。

【0025】ここで、情報分布空間生成部30が一実施形態として用いるKohonenの自己組織化マップによる情報分布空間の生成処理を詳しく説明しておく。

【0026】自己組織化マップでは、類似度を内積として以下のように実現している。

【0027】（1）図2（a）は学習格子と学習範囲の概念を模式的に示す図である。いま、図2（a）のような $L \times L$ の正方格子を持つ2次元空間を考える。ここで、 L は各格子に1つずつ n 個の情報を配置するために n の二乗根を超える最小の整数とする。各格子点（ i ,

$$m_{i,j,k}(t+1) = m_{i,j,k}(t) + c(t) e^{-ar^2} (x_k - m_{i,j,k}(t))$$

r は格子点間のユークリッド距離で、 a は1より小さい係数である。

$$c(t) = e^{-a} \quad : t < G \times n \text{ (図2 (b) の初期学習の範囲)}$$

$$c(t) = e^{-a} \times G \times n / t \quad : t \geq G \times n \text{ (図2 (b) の調整学習の範囲)}$$

【0034】（5） $m_{i,j,k}(t+1)$ を正規化し、それをその格子点の新しい特徴ベクトルとする。

【0035】（6）上記（3）から（5）を対象情報の集合単位（ n ）毎に各データの特徴ベクトルを入力データとして実行し、各特徴ベクトルに最も類似度の高い格子点のマップベクトルとその特徴ベクトルの類似度の最小値が一定の値以上になるか、その最小値の変化が一定の値以下になるまで繰り返す。例えば、収束条件として前回との最小値の差が有効数字の上3桁が同一になるまで実施する。

【0036】（7）収束した2次元空間の格子点のマップベクトルを使い、各情報の特徴ベクトルに最も類似度の高いマップベクトルを持つ格子点にその情報を配置する。

【0037】次に、応用形態として、自己組織化マップにおける前記情報分布空間への入力情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を制限

j ）に対応して、マップベクトル（ $m_{i,j,k}$ ）を学習させることを考える。 k は選択された特徴表現を表わし、入力の特徴ベクトルと同じ要素数のベクトルを表す。

【0028】（2）各情報の特徴ベクトル（ x_i ）は（数1）の形に正規化されているものとする。

【0029】

【数1】

$$\sqrt{\sum x_i^2} = 1$$

【0030】（3）類似度 $s(x, y)$ は、内積（数2）で表現され、この類似度により特徴ベクトル（ x_i ）に最も類似度の高い格子点（ $m_{i,j,k}$ ）を選択する。

【0031】

【数2】

$$\sum x_i y_i$$

【0032】（4）選択した最適な格子点を中心として、学習回数と学習範囲の関係を図2（b）のように近似し、これで決まる半径 h の範囲内の格子点のマップベクトルを（数3）式で更新していく。図2（b）の G は初期学習の範囲の終了点を表わし、入力データ数 n を単位とした学習回数である。 T は調整学習の緩やかに変化させる範囲の開始点であり、 n を単位とした学習回数である。

【0033】

【数3】

する方法を説明する。例えば、上記実装法における

（3）の処理の中で、マップ全体に対して類似度を計算することに代え、図2（c）のグラフによって決まる半径内の格子点のみを探索範囲の対象として類似度を計算することにより学習位置を決定する。図2（c）は、学習回数に応じた探索半径の制御の様子を模式的に示した図である。学習対象となる情報全体を1セットととして学習を繰り返すので、学習位置を決める対象情報に対して、前回の学習位置を中心とした最大探索半径が学習回数とともに変化する様子が示されている。初回の場合には、前回の学習位置がないため、全格子に対する探索を実施する。図2（c）では3セット回（ $3n$ ）まで全格子の探索を実施し、4回目は半径を減少させ、5回目から n で決まる半径（ $f(n)$ ）に探索範囲を絞っている。この $f(n)$ の関数の例としては、 $n/200+2$ がある。

【0038】このように探索範囲を情報分布空間の一辺

の格子数 L により決まる一定半径に制限するので情報分布空間の生成が短時間で効率良く実行することができ、また、上記に示したように調整学習において探索範囲を制限しても当初のマップ全体を探索範囲とする初期学習と組み合わせることにより一定レベルの特徴ベクトルの分布を持つ情報分布空間を得ることができる。

【0039】情報分布処理部40は、入力部10から入力された全情報をそれら情報を持つ特徴ベクトルの類似度により情報分布空間生成部30が生成した情報分布空間に対して分布・配置処理を実行する部分である。情報分布処理部40は、次の情報分布空間拡大機能と情報配

$$M = \lceil \sqrt[N]{N} \rceil \quad (N \text{ の } 2 \text{ 乗根を越える最小の整数})$$

【0042】情報分布空間の一辺を L から M へ拡張する処理の一例は以下のようである。

【0043】(1) x 方向の L 点の特徴ベクトルから、 M 点の特徴ベクトルを算出マップ上の y 方向の格子点毎に距離の比例配分により算出する。例えば、マップの i

$$v_k'(j) = (v_k(i) \times r_{i,j} + v_k(i+1) \times r_{j,j+1}) / (r_{i,j} + r_{j,j+1})$$

ここで $v_k'(j)$ は拡張後の j 点における特徴ベクトル、

$v_k(i)$ は拡張前の i 点における特徴ベクトル、

r は x 軸上の位置

【0045】(2) y 方向の算出では、拡大マップ上 x 方向の格子点毎に、(1)と同様な比例配分の計算により特徴ベクトルを求める。

【0046】情報配置機能は、必要な上記情報分布空間拡大機能による情報分布空間の拡大後、全入力情報ごとにその特徴ベクトルとマップの格子点の特徴ベクトルとの間の類似度が最大になる格子点を配置位置として求め、該格子点に該情報を配置する機能である。

【0047】表示部50は、利用者に対して得られた情報分布を表示する部分である。例えば、コンピュータの表示処理関連の機能とモニタを備えたものである。

【0048】上記構成の本実施形態1の情報分布装置の処理の流れをフローチャートを参照しつつ具体例を交えて説明する。

【0049】図3は実施形態1の情報分布装置の処理の流れを示すフローチャートである。この例において、入力された情報が、データベース検索アプリケーションのキーワード検索などにより10,500件ヒットした画像情報であるとする。特徴ベクトルとして色相ベクトルが与えられたものとする。また、抽出する学習対象情報数を1,089件(33×33)とする。

【0050】まず、入力部10を介して分布配置する情報の集合および特徴ベクトルの集合が入力される(ステップS301)。ここでは、10,500件の画像情報の集合とその特徴ベクトルの集合が入力される。入力された情報の集合と特徴ベクトルの集合は抽出部20およ

び機能部を備えている。

【0040】情報分布空間拡大機能とは、情報分布空間生成部30により生成された情報分布空間が持つ格子点より、配置する全入力情報の方が大きい場合、全入力情報が適度に一覧性をもって表示できるように情報分布空間を拡大する機能である。例えば、全体の入力情報数を N とすると、拡大されたマップの一辺は(数4)に示す M とすれば良い。

【0041】

【数4】

番目と $i+1$ 番目の間に拡大マップの j 点があるとする、 j 点の特徴ベクトルは以下の(数5)のように算出できる。

【0044】

【数5】

び情報分布処理部40に渡される。

【0051】次に、抽出部20により、情報の持つ特徴ベクトルが単位ベクトルとゼロベクトルであるものに低い優先順位をつける(ステップS302)。つまり、情報の持つ特徴ベクトルが単位ベクトルとゼロベクトルであるものの抽出優先順位を低くし、単位ベクトルとゼロベクトル以外の特徴ベクトルを持つ情報を優先的に抽出するようにする。

【0052】次に、抽出部20は、入力された情報数が所定数より大きい数であるかを調べ、大きい場合、所定数の情報を抽出する(ステップS303)。ここでは所定数1,089件より全入力情報数が大きいので、1,089件の情報が学習対象情報として抽出される。なお、抽出するアルゴリズムとしては様々な想定できるが、この例では、単位ベクトルとゼロベクトル以外の特徴ベクトルを持つ情報からランダム抽出するものとする。抽出部20は抽出した学習対象情報を情報分布空間生成部30に渡す。

【0053】次に、情報分布空間生成部30により、自己組織化マップにより学習対象情報を用いて情報分布空間を生成する(ステップS304)。生成した結果は特徴ベクトルが対応づけられた $L \times L$ の格子を持つ情報分布空間になる。この例では、 L は33以下の自然数となる。

【0054】情報分布空間生成部30は、ステップS304の情報分布空間生成処理の中で、応用機能として、

上記したような最適位置探索範囲制御機能を用いる（ステップS304-1）。つまり、特徴ベクトルの対応する格子上の学習位置を決めるため、入力の特徴ベクトルと格子上の特徴ベクトルの類似度が最も高い位置を探索する範囲を制御する。ここでは、自己組織化マップにおける情報分布空間への入力情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じた制御を実行する。この例では4回目以降の最適位置の探索範囲を $N/200+2$ （格子間の距離を1とする）として実行する。

【0055】情報分布空間生成部30は、生成した情報分布空間を情報分布処理部40に渡す。この例では、 33×33 の格子を持つ情報分布空間が生成されている。

【0056】次に、情報分布処理部40は、上記した情報分布空間拡大機能により情報分布空間を全入力情報数に対応する空間に拡大する（ステップS305）。この例では、 33×33 の格子を持つ情報分布空間を 103×103 の格子を持つ情報分布空間に拡大処理する。

【0057】次に、情報分布処理部40は、上記した情報配置機能を用いて、拡大した情報分布空間に対して全入力情報を配置する（ステップS306）。この例では、 103×103 の格子を持つ情報分布空間に対して10,500件の情報ごとに、その特徴ベクトルとマップの格子点の特徴ベクトルとの間の類似度が最大になる格子点を配置位置として求められ、配置される。

【0058】情報分布処理部40は、生成結果として分布配置済みのデータを出力し（ステップS307）、必要に応じて表示部50において結果表示を行う。

【0059】ここで、情報分布空間の例を示す。まず、参考として、上記ステップS305の情報分布空間拡大機能による情報分布空間の拡大処理がされなかった場合の情報分布空間と配置された1,050件の学習対象情報の表示例を図4に示す。図4は 33×33 の格子点を持つ情報分布空間であり、1,050の学習対象情報である画像情報が色相を特徴ベクトルとして適度に分類されて表示されている。図5は、上記ステップS305の情報分布空間拡大機能による情報分布空間の拡大処理実行後の情報分布空間と配置された10,500件の全入力情報の表示例である。図5は 65×65 の格子点を持つ情報分布空間であり、10,500の全入力情報である画像情報が色相を特徴ベクトルとして適度に分類されて表示されている。上記図4の状態から図5の状態を得るために必要な計算時間は、単純なマップの拡大処理計算と各情報をマッピングしてゆくのみであるので短時間となる。

【0060】以上が、本実施形態1の情報分布装置の処理の流れである。

【0061】以上、本実施形態1の情報分布装置および情報分布方法によれば、全入力情報から学習対象となる情報を抽出して情報分布空間を生成するので、処理時間

の効率化と分布結果の品質維持の両方を同時に実現することができる。また、自己組織化マップにおける情報分布空間への入力情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じた制御を実行するのでさらに処理時間を効率的に短縮することができる。

【0062】（実施形態2）実施形態2の情報分布装置は、実施形態1で説明した情報分布処理部がさらに再配置処理部を備え、全入力情報を情報分布空間へ分布した場合に同じ位置に重複配置される情報があれば、再配置処理部により、重複配置された情報のうち再配置する情報を選んで再配置し、情報の一覧表示を見易くするものである。つまり、格子位置に対応する特徴ベクトルが適度に分布している情報分布空間を生成しても入力される情報に如何によって、また、情報数が大きくなると同じ格子点に複数の情報が重複配置されてしまう場合が有り得る。この場合に重複配置したままでは情報の一覧性を損なうので再配置するものである。

【0063】図6は、本実施形態2の情報分布装置の構成例を示す図である。図6において、入力部10、抽出部20、情報分布空間生成部30、表示部50は実施形態1で示した図1の各要素と同様のものであり、ここでの説明は適宜省略する。

【0064】情報分布処理部40aは、再配置処理部41を備えている。再配置処理部41は、全入力情報を情報分布空間へ分布した場合に重複配置される情報があれば、重複配置された情報のうち再配置する情報を選んで再配置処理する部分である。再配置処理部41は以下の再配置情報選択機能と再配置機能を備えている。再配置情報選択機能は、実施形態1に示した処理により全入力情報を特徴ベクトルの類似度により情報分布空間へ分布した場合に同じ位置に重複配置されている情報を検索し、それら情報から再配置処理する情報を選ぶ。図7は、一つの情報分布空間の格子に複数の情報が重複配置されている様子とそれら情報を再配置する概念を模式的に示した図である。説明の便宜上、9枚の画像情報が一つの情報分布空間の格子に重複配置されている単純な例を示した。図7(a)に示すように、9枚の画像情報が一つの情報分布空間の格子に重複配置されている。ここで、再配置情報選択機能の例として、当該位置に対する特徴ベクトルと重複配置されている各情報が持つ特徴ベクトルとの類似度を比較し、もっとも類似度が高い情報をその位置に残し、他の情報を再配置処理に供する。このように再配置する情報を選択すれば、重複配置されていた位置にはもっとも類似度の高い情報の配置が維持される。図7(b)は再配置した様子の一例であるが、下記に示すように再配置機能には複数通りの処理方法がある。

【0065】第1の処理は、再配置する情報を重複配置位置周囲の一定範囲で情報が配置されていない位置に再

配置する処理である。つまり、一定範囲に再分布するため、情報配置の違和感を少なくし、かつ、一覧性を上げることができる。近傍の一定範囲で情報が配置されていない位置が複数ある場合、例えば次の基準で再配置位置を決める。図8は重複配置位置周囲の一定範囲で情報を再配置する優先順位を示す図である。この例では重複配置位置から反時計周りの螺旋状に優先順位を与えている。図7(b)はこの図8の基準に沿って再配置した例となっている。なお、元の配置位置から再配置を許す位置までの距離、つまり、近傍と扱う範囲については、一定の制限を設けることが好ましい。あまりに離れた位置に再配置すると情報配置の違和感が大きくなるからである。

【0066】第2の処理は、再配置する情報を重複配置位置周囲で情報が配置されていない位置であってその位置に対する特徴ベクトルと再配置する情報が持つ特徴ベクトルとの類似度が一定範囲にある位置に再配置する処理がある。つまり、近傍で情報が配置されていない位置にはそれぞれ特徴ベクトルが対応しており、再配置する情報の特徴ベクトルとの類似度合いはまちまちであるので、類似度合いの大小で直接比較し、もっとも類似度合いが高い位置に再配置するものである。

【0067】第3の処理は、再配置する情報を重複配置位置周囲の一定範囲でかつ情報が配置されていない位置であってその位置に対する特徴ベクトルと再配置する情報が持つ特徴ベクトルとの類似度が一定範囲にある位置に再配置する処理がある。つまり、近傍の一定範囲に再分布して情報配置の違和感を少なくするという第1の処理の利点と再配置位置の特徴ベクトルと再配置する情報の特徴ベクトルとの類似度合いを高く保つという第2の処理の利点の双方を兼ね備えた処理とするものである。

【0068】第4の処理は、再配置する情報を重複配置位置周囲で情報が配置されていない位置であってその位置に対する特徴ベクトルと再配置する情報が持つ特徴ベクトルとの類似度と、重複配置の位置と情報が配置されていない位置との距離との2つのパラメータを持つ関数とし、関数値を評価して再配置する位置を決める処理である。これは、第3の処理において考慮した2つのパラメータである、再配置する位置の距離と特徴ベクトルの類似度合いを関数として定義し、きめ細やかに再配置処理の調整を行うものである。

【0069】第5の処理は、再配置する情報を重複配置位置周囲で情報が配置されている位置を選び、選択位置に対する特徴ベクトルと再配置する情報が持つ特徴ベクトルとの類似度が選択位置に対する特徴ベクトルと既に選択位置に配置されている情報が持つ特徴ベクトルとの類似度より大きい場合、再配置する情報を選択位置の前面に配置する処理である。つまり、重複配置された情報の中には、複数の格子点の特徴ベクトルと良く類似しており、もっとも類似する格子点に配置されたものの、よ

り類似度合いが高い他の情報の存在により再配置に供される場合が有り得る。この場合、2番目に類似度合いが高い格子点には既に他の情報が配置済みである場合も想定できる。この場合、既に配置済みの情報より、当該再配置に供される情報の方がその格子点の特徴ベクトルとの類似度合いが高い場合も想定できる。第5の処理はかかる場合に、既に配置済みの情報が存在してももっとも類似度合いが高い情報が格子点に配置されるように配置を柔軟に見直して再配置処理を実行するものである。第5の処理によれば、一度配置された情報も含め、情報全体としてもっとも格子点の特徴ベクトルとの類似度合いが高くなるように再配置することができる。ただし、第5の処理により配置を見直す範囲も一定の制限を設けておくことが好ましい。再配置処理の連鎖が広がるおそれもあるからである。

【0070】再配置処理部41は、情報分布空間における重複配置された情報の再配置処理として、上記の第1の処理から第5の処理のいずれかの1つの処理、または、それらの組み合わせの処理を用いることができるものである。組み合わせる用いる場合、例えば、最初、第1の処理と第2の処理の組み合わせである第3の処理により再配置を実行し、未だ重複配置が残り、近傍に情報が未配置の格子点がなくなった場合に第5の処理を実行するといった組み合わせがある。

【0071】次に、上記構成の本実施形態2の情報分布装置の処理の流れをフローチャートを参照しつつ具体例を交えて説明する。

【0072】図9は実施形態2の情報分布装置の処理の流れを示すフローチャートである。図9のステップS901からステップS905は、実施形態1で説明した図3のフローチャートのステップS301からステップS305と同様であり、ここでの説明は適宜省略する。

【0073】ステップS906において、情報分布処理部40aは、拡大した情報分布空間に対して全入力情報を配置している。ステップS907において、情報分布処理部40aの再配置処理部41は、情報分布空間において重複配置されている情報の有無を調べる(ステップS907)。

【0074】重複配置されている情報がある場合(ステップS907:Y)、再配置処理を実行する(ステップS908)。例えば、再配置処理部41は再配置処理として、選択された上記の第1の処理から第5の処理のいずれか1つの処理、または、それらの組み合わせの処理を実行するものとする。

【0075】再配置処理(ステップS908)を実行後、ステップS907に戻り重複配置されている情報の有無を調べ、重複配置されている情報がある場合(ステップS907:Y)、再配置処理(ステップS908)を繰り返して実行する。

【0076】重複配置されている情報がない場合(ステ

ップS907:N)、情報分布処理部40は、生成結果として分布配置済みのデータを出力し(ステップS909)、必要に応じて表示部50において結果表示を行う。

【0077】再配置処理の例を示す。ここではテキストの配置例とその重複配置の再配置例を示す。図10および図11は、再配置する前の様子であり、写真データのタイトルと写真家の名前のテキストデータを情報とし、そのテキストデータに対して25のキーワードを特徴ベクトルとして与えたものを入力し、それらを情報分布空間に配置した表示例である。このような例では、テキスト内容が単純なものであるため、一つの情報に対応するキーワードは少なく、重複配置されるものが多い。重複配置された情報がZ軸方向(紙面垂直奥方向)に螺旋状に配置されている様子が分かる。なお、図11は図10の中央部分を拡大して表示した図である。図12は図10の状態に対して再配置処理を実行したもので、図13は図11に相当する部分を拡大したものとなっている。このように写真家「芳賀」のキーワードで配置された情報が周囲に再配置されている様子が分かる。

【0078】以上、実施形態2の情報分布装置および情報分布方法によれば、再配置処理部を備え、全入力情報を情報分布空間へ分布した場合に同じ位置に重複配置される情報があれば、再配置処理部により、重複配置された情報のうち再配置する情報を選んで再配置し、情報の一覧表示を見易くすることが可能となる。

【0079】(実施形態3)本発明の情報分布装置および情報分布方法は、上記に説明した構成を実現する処理ステップを記述したプログラムをコンピュータ読み取り可能な記録媒体に記録して提供することにより、各種コンピュータを用いて構築できる。本発明の情報分布装置および情報分布方法を実現する処理ステップを備えたプログラムを記録した記録媒体は、図14に図示した記録媒体の例に示すように、CD-ROM1002やフレキシブルディスク1003等の可搬型記録媒体1001だけでなく、フィールド上にある記録装置内の記録媒体1000や、コンピュータのハードディスクやRAM等の記録媒体1005のいずれであっても良く、プログラム実行時には、プログラムはコンピュータ1004上にローディングされ、主メモリ上で実行される。

【0080】さらに、ソースプログラムをコンパイルしたもののみならず、フィールドを介してクライアントコンピュータに中間言語形式のアプレットを送信し、クライアントコンピュータ上でインタープリタ実行して動作する構成であっても良い。

【0081】本発明の情報分布装置および情報分布方法に関し、以下の項を開示する。

【0082】(付記1)情報の集合とその情報に対応する特徴ベクトルの集合を入力する入力部と、前記入力された情報の数が所定数より多い場合、その中から所定数

の情報を抽出する抽出部と、前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布する情報分布空間を生成する情報分布空間生成部と、前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布する情報分布処理部を備えたことを特徴とする情報分布装置。

【0083】(付記2)前記情報分布空間への特徴ベクトルの類似度を基にした情報の分布処理として、自己組織化マップによる分布処理を用い、前記情報分布空間生成部が、前記自己組織化マップにおける前記情報分布空間への学習対象情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じて制限する上記付記1に記載の情報分布装置。

【0084】(付記3)前記抽出部が、入力された情報の中から所定数の情報を抽出するにあたり、特徴ベクトルが単位ベクトルである情報以外の情報を優先して抽出する上記付記1に記載の情報分布装置。

【0085】(付記4)前記情報分布処理部が再配置処理部を備え、前記再配置処理部は、前記入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選び、該再配置する情報を該重複配置位置周囲の一定範囲で情報が配置されていない位置に再配置する上記付記1に記載の情報分布装置。

【0086】(付記5)前記情報分布処理部が再配置処理部を備え、前記再配置処理部は、前記入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選び、該再配置する情報を該重複配置位置周囲で情報が配置されていない位置であってその位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度が一定範囲にある位置に再配置する上記付記1に記載の情報分布装置。

【0087】(付記6)前記情報分布処理部が再配置処理部を備え、前記再配置処理部は、前記入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選び、該再配置する情報を該重複配置位置周囲の一定範囲でかつ情報が配置されていない位置であってその位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度が一定範囲にある位置に再配置する上記付記1に記載の情報分布装置。

【0088】(付記7)前記情報分布処理部が再配置処理部を備え、前記再配置処理部は、前記入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選び、該再配置する情報を該重複配置位置周囲で情報が配置されていない位置であって、その位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベ

クトルとの類似度と、前記重複配置の位置と前記情報が配置されていない位置との距離との2つのパラメータを持つ関数とし、前記関数値を評価して再配置する位置を決める上記付記1に記載の情報分布装置。

【0089】（付記8）前記情報分布処理部が再配置処理部を備え、前記再配置処理部は、前記入力された全情報を特徴ベクトルの類似度により前記情報分布空間へ分布した場合に同じ位置に重複配置される情報のうち再配置する情報を選び、該再配置する情報を該重複配置位置周囲で情報が配置されている位置を選び、該選択位置に対する特徴ベクトルと該再配置する情報が持つ特徴ベクトルとの類似度が該選択位置に対する特徴ベクトルと既に該選択位置に配置されている情報が持つ特徴ベクトルとの類似度より大きい場合、該再配置する情報を該選択位置の前面に配置する上記付記1に記載の情報分布装置。

【0090】（付記9）情報の集合とその情報に対応する特徴ベクトルの集合を入力し、前記入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出し、前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布する情報分布空間を生成し、前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布することを特徴とする情報分布方法。

【0091】（付記10）情報の集合をその特徴ベクトルを基に分布する情報分布装置を構成する処理プログラムを記録したコンピュータ読み取り可能な記録媒体であって、情報の集合とその情報に対応する特徴ベクトルの集合を入力する処理ステップと、前記入力された情報の数が所定数より多い場合、その中から所定数の情報を抽出する処理ステップと、前記抽出された情報を学習対象情報として用いて学習し、該学習対象情報が持つ特徴ベクトルの類似度により該学習対象情報を分布する情報分布空間を生成する処理ステップと、前記入力された全情報をそれら情報が持つ特徴ベクトルの類似度により前記情報分布空間に分布する処理ステップとを備えた処理プログラムを記録したことを特徴とする記録媒体。

【0092】

【発明の効果】本発明の情報分布装置および情報分布方法によれば、全入力情報から学習対象となる情報を抽出して情報分布空間を生成するので、処理時間の効率化と分布結果の品質維持の両方を同時に実現することができる。また、自己組織化マップにおける情報分布空間への入力情報の特徴ベクトルの分布最適位置をマップから探す場合の探索範囲を学習回数に応じた制御を実行するのでさらに処理時間を効率的に短縮することができる。

【0093】また、本発明の情報分布装置および情報分布方法によれば、再配置処理部を備え、全入力情報を情報分布空間へ分布した場合に同じ位置に重複配置される

情報があれば、再配置処理部により、重複配置された情報のうち再配置する情報を選んで再配置し、情報の一覧表示を見易くすることが可能となる。

【図面の簡単な説明】

【図1】 本発明の実施形態1の情報分布装置の構成例を示す図

【図2】 （a）は学習格子と学習範囲の概念を模式的に示す図、（b）は学習回数と学習範囲の関係を示す図、（c）は学習回数に応じた探索半径の制御の様子を模式的に示した図

【図3】 実施形態1の情報分布装置の処理の流れを示すフローチャート

【図4】 情報分布空間の拡大処理実行前の情報分布空間と配置された1,050件の学習対象情報の表示例を示す図

【図5】 情報分布空間の拡大処理実行後の情報分布空間と配置された10,500件の全入力情報の表示例を示す図

【図6】 本発明の実施形態2の情報分布装置の構成例を示す図

【図7】 一つの情報分布空間の格子に複数の情報が重複配置されている様子とそれら情報を再配置する概念を模式的に示した図

【図8】 重複配置位置周囲の一定範囲で情報を再配置する優先順位を示す図

【図9】 実施形態2の情報分布装置の処理の流れを示すフローチャート

【図10】 再配置する前の、写真データのタイトルと写真家の名前のテキストデータを情報とし情報分布空間に配置した表示例を示す図

【図11】 図10の中央部分を拡大した図

【図12】 図10の状態に対して再配置処理を実行した様子を示す図

【図13】 図12の中央部分を拡大した図であり、図11に相当する部分を示す図

【図14】 本発明の情報分布装置および情報分布方法を実現する処理ステップを記述したプログラムを記録した記録媒体の例を示す図

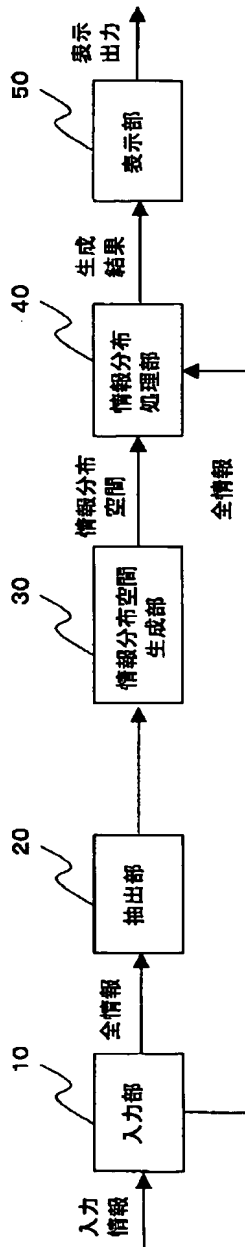
【符号の説明】

- 10 入力部
- 20 抽出部
- 30 情報分布空間生成部
- 40, 40a 情報分布処理部
- 41 再配置処理部
- 50 表示部
- 1000 記録装置内の記録媒体
- 1001 可搬型記録媒体
- 1002 CD-ROM
- 1003 フレキシブルディスク
- 1004 コンピュータ

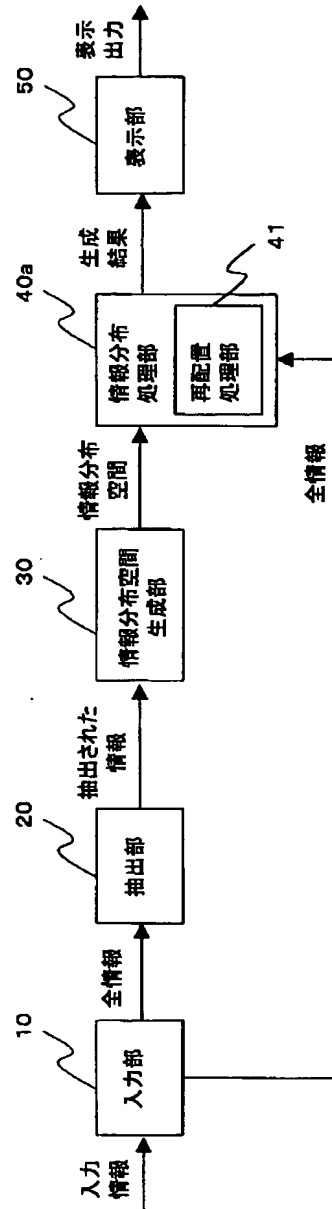
1005 コンピュータのハードディスクやRAM等の

記録媒体

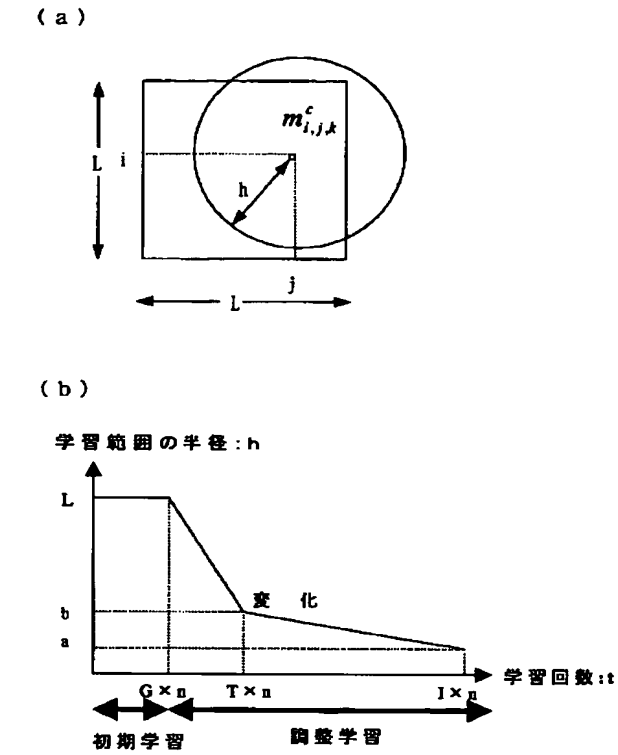
【図1】



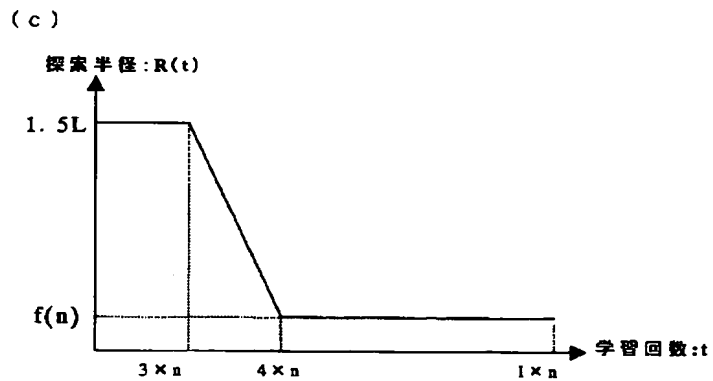
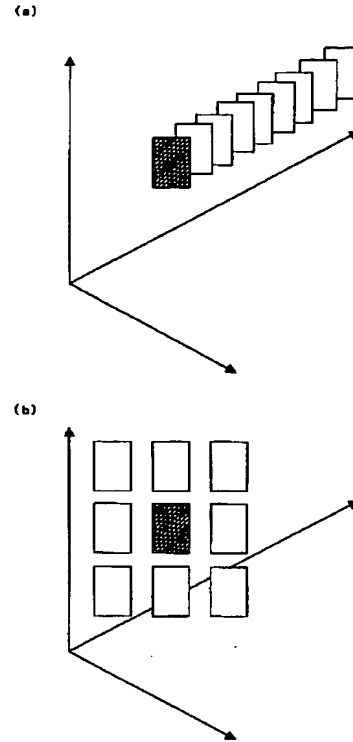
【図6】



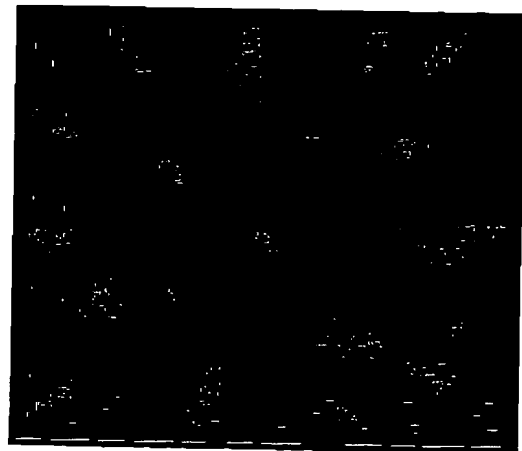
【図2】



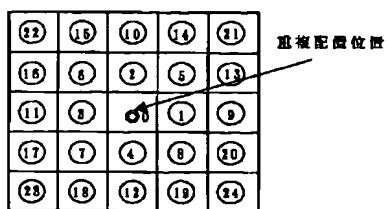
【図7】



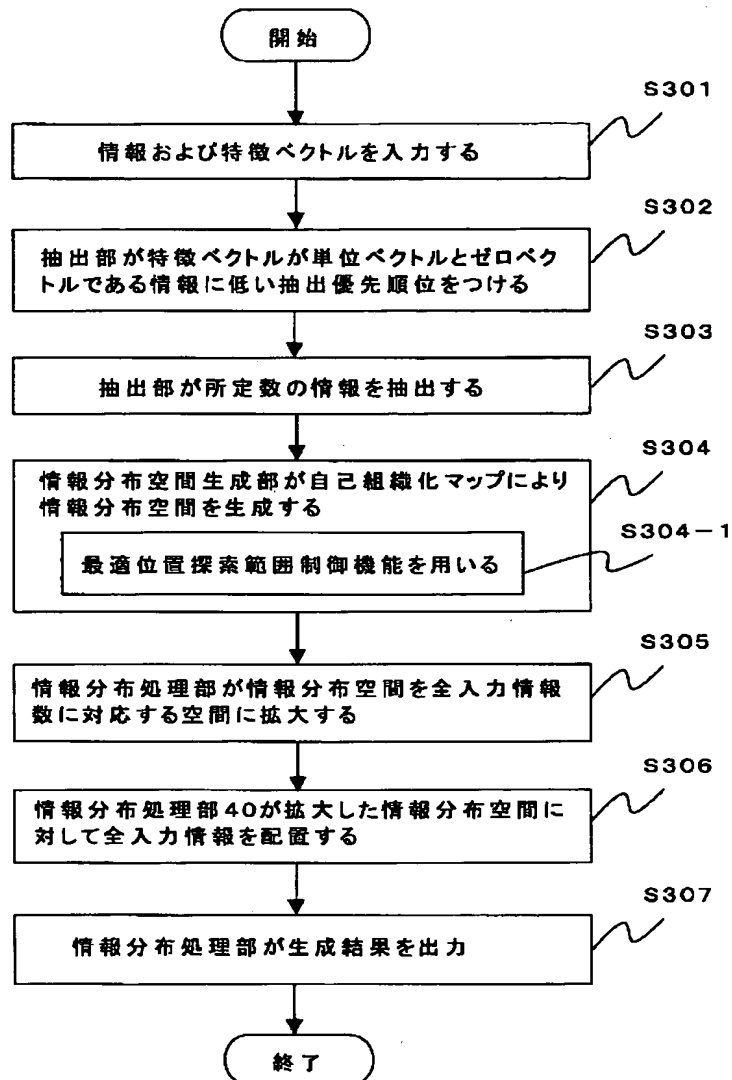
【図10】



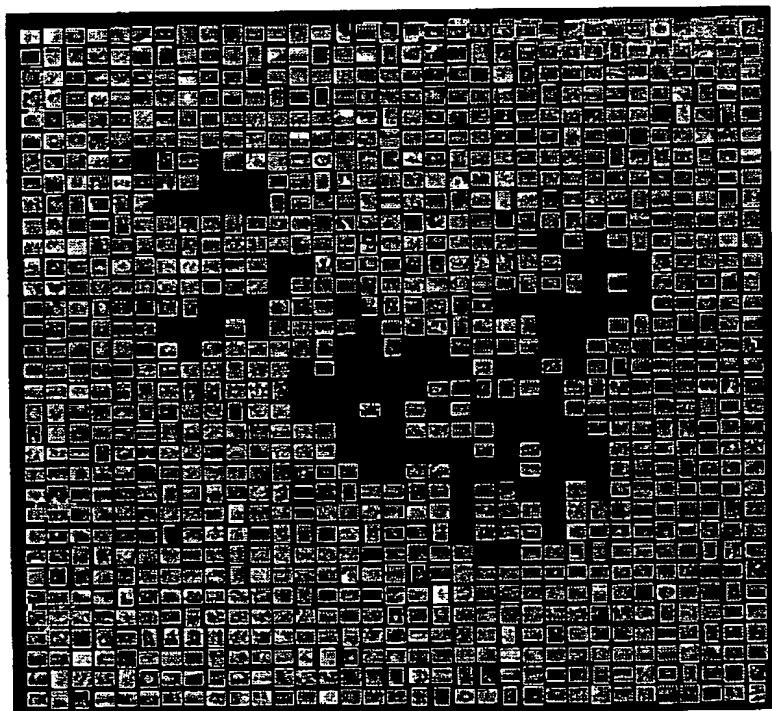
【図8】



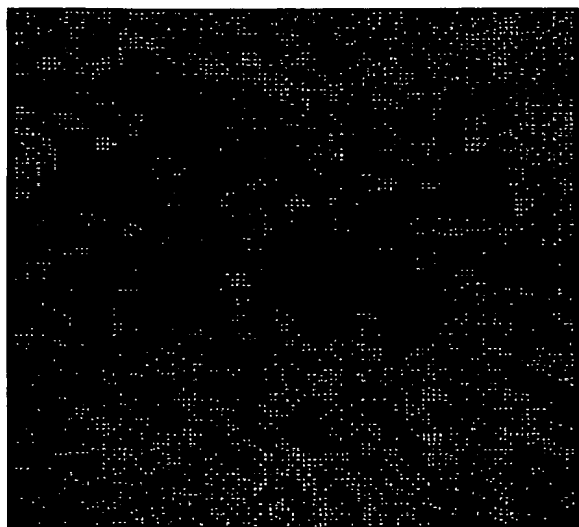
【図3】



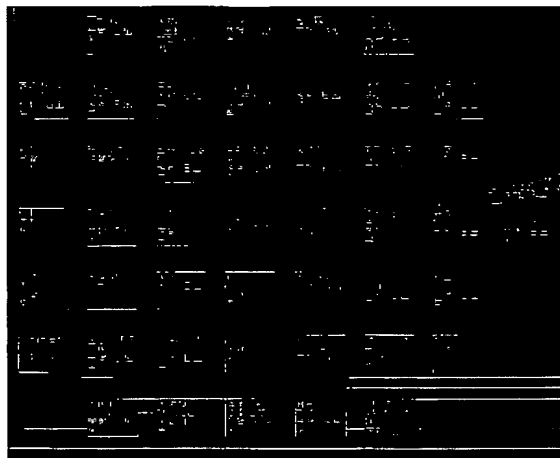
【図4】



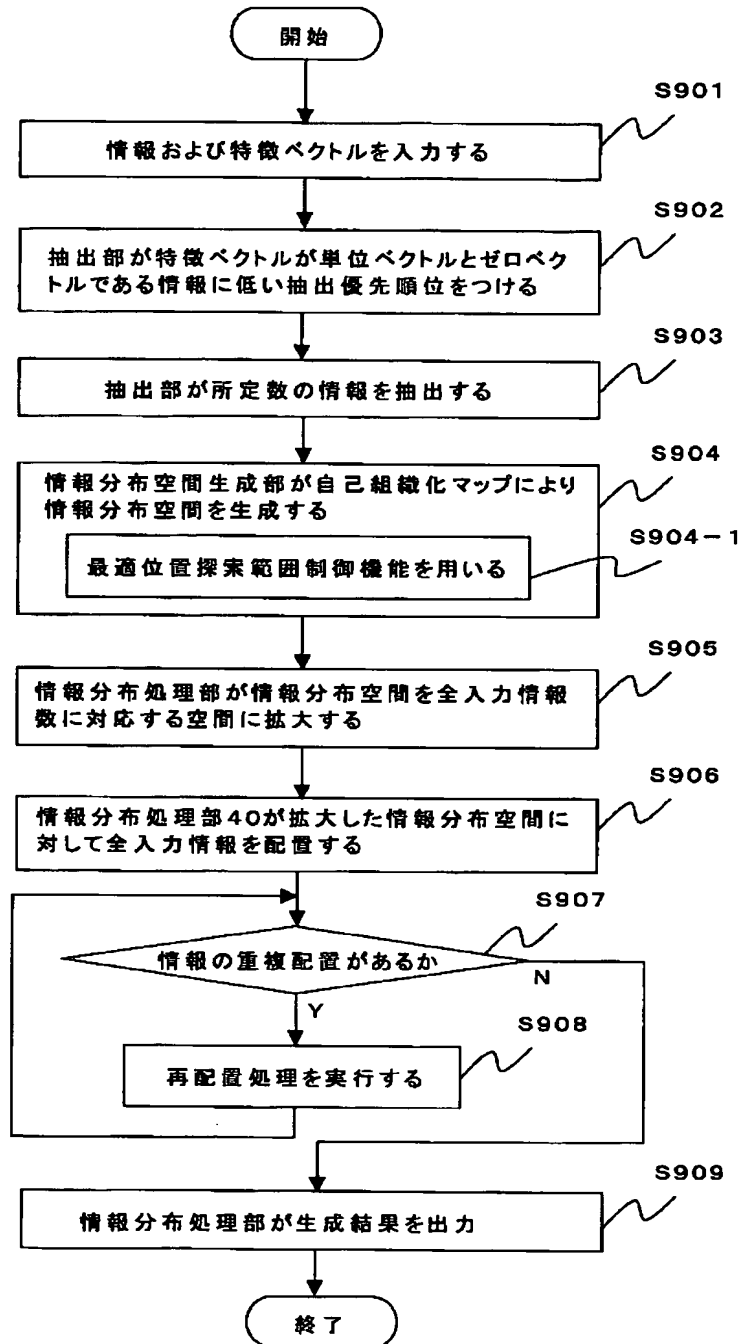
【図5】



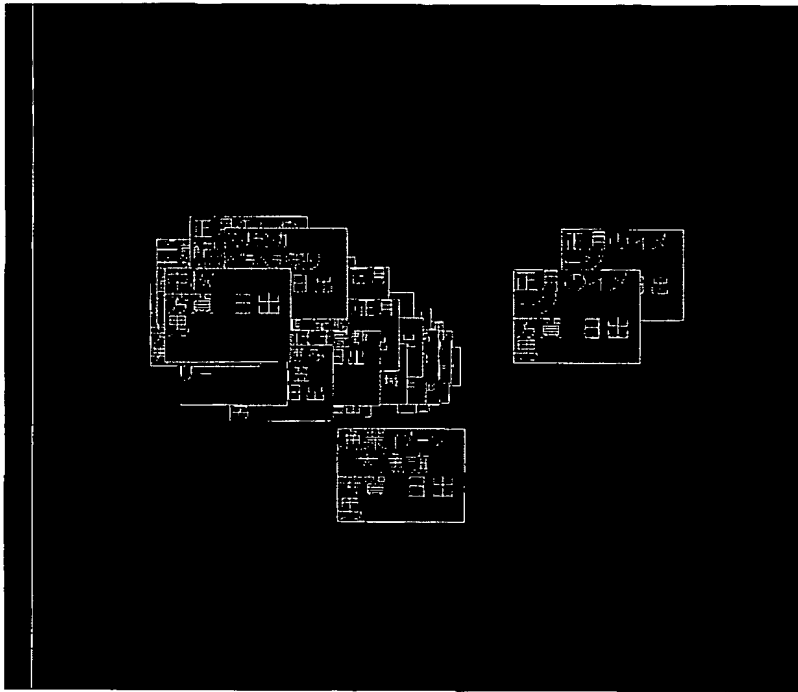
【図13】



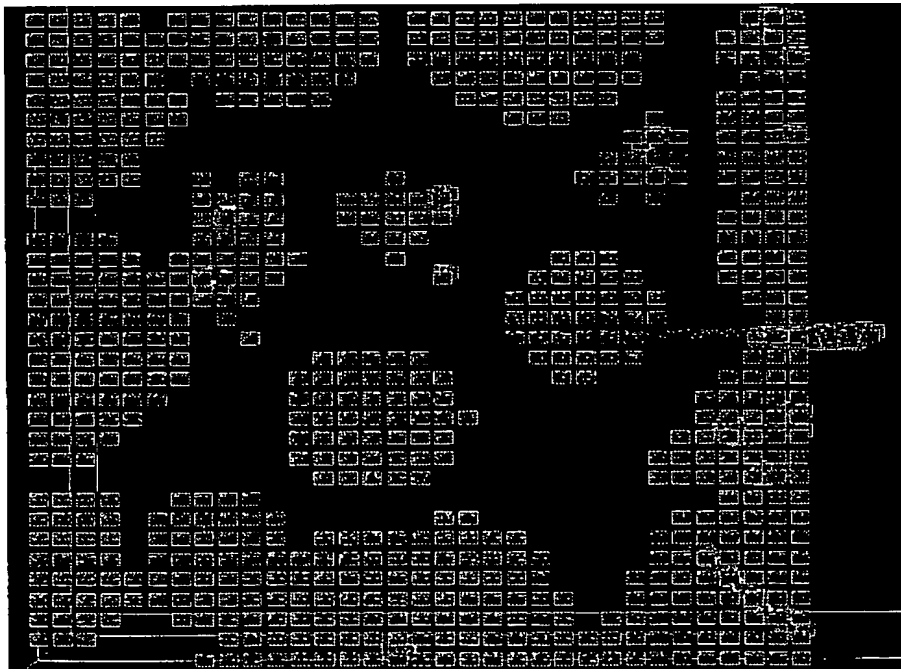
【図9】



【図11】



【図12】



【図14】

